JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 162, 196–209 (1991)

How to Regularize a Difference of Convex Functions

J.-B. HIRIART-URRUTY

Laboratoire d'Analyse Numérique, U.F.R. Mathématiques, Informatique, Gestion, Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse Cedex, France

Submitted by Richard M. Aron

Received January 29, 1990

Given a nonconvex function f defined as the difference of two convex functions g and h (f is a so-called d.c. function), we study the regularized (or smoothed) version $f_r = g \,\Box\, r/2 \,\|\cdot\|^2 - h \,\Box\, r/2 \,\|\cdot\|^2$ of f obtained by performing the infimal convolution of both component functions g and h by the same kernel function $r/2 \,\|\cdot\|^2$. Critical points of f_r and f are compared and the behavior of critical points of f_r as $r \to +\infty$ is considered. To a great extent the nice properties of the regularization process $\varphi \to \varphi \,\Box\, r/2 \,\|\cdot\|^2$ when applied to convex functions φ are preserved for the process $f \to f_r$ when performed on d.c. functions f.

I. INTRODUCTION

Let $(H, \langle \cdot, \cdot \rangle)$ be a real Hilbert space; we denote by $\|\cdot\|$ the norm associated with the inner product $\langle \cdot, \cdot \rangle$, and the topological dual space H' is identified with H. Throughout we use the following notations, which are of a common use in the context of convex analysis and optimization:

- $\Gamma_0(H)$ is the set of *lower-semicontinuous proper convex functions* from H into $(-\infty, +\infty]$ (a function $\varphi: H \to (-\infty, +\infty]$ is said to be proper if φ is not identically equal to $+\infty$ and if $\varphi(x) > -\infty$ for all $x \in H$).

- Given $\varphi \in \Gamma_0(H)$, dom φ (*domain of* φ) is the set of x at which φ is finite. For $\varepsilon \geqslant 0$, the ε-*subdifferential* of φ at x_0 is defined as the set of vectors $x^* \in H$ satisfying

$$\varphi(x) \geqslant \varphi(x_0) + \langle x^*, x - x_0 \rangle - \varepsilon \qquad \text{for all } x \in H \tag{1.1}$$

The set of such vectors, denoted by $\partial_\varepsilon \varphi(x_0)$, is a closed convex set in H which reduces to the usual (or "exact") subdifferential $\partial\varphi(x_0)$ when $\varepsilon = 0$.

- The *conjugate* (*or polar*) *function* of $\varphi \in \Gamma_0(H)$ is the function $\varphi^* : H \to (-\infty, +\infty]$ defined by

$$\forall x^* \in H, \qquad \varphi^*(x^*) = \sup_{x \in H} \{\langle x, x^* \rangle - \varphi(x)\}. \tag{1.2}$$

0022-247X/91 $3.00

Then $\varphi^* \in \Gamma_0(H)$ and a fundamental property of the conjugacy operation is that $(\varphi^*)^* = \varphi$.

A nice way of regularizing $\varphi \in \Gamma_0(H)$ is to perform the *infimal convolution* of φ with some kernel function $\theta \in \Gamma_0(H)$; the resulting function, denoted as $\varphi \,\square\, \theta$ assigns to $x \in H$ the value

$$(\varphi \,\square\, \theta)(x) = \inf_{\substack{x_1, x_2 \in H \\ x_1 + x_2 = x}} \{\varphi(x_1) + \theta(x_2)\}.$$

As a general rule $(\varphi \,\square\, \theta)^* = \varphi^* + \theta^*$, and under some additional assumptions on φ and θ, $(\varphi^* + \theta^*)^* = \varphi \,\square\, \theta$ (cf. [1, 12] for instance). Among all the possible kernel functions $\theta \in \Gamma_0(H)$, there is one of a particular interest, that is $\theta_0 = \frac{1}{2} \|\cdot\|^2$. It turns out that θ_0 is the only function on H for which $\theta_0^* = \theta_0$, and, clearly, $\nabla\theta_0(x) = x$ for all $x \in H$. Performing the infimal convolution of $\varphi \in \Gamma_0(H)$ with $r\theta_0$, $r > 0$, gives rise to a particularly interesting regularized version $\varphi_r = \varphi \,\square\, r\theta_0$ of φ. The regularization process $\varphi \to \varphi_r$, initiated by Moreau [13], is now well understood and widely displayed in the field of convex analysis and optimization, even in textbooks [1, 2, 6]). Moreover it contains the theoretical roots for algorithmic procedures like the so-called "proximal methods" in convex minimization; see [19] for a recent survey on the subject. For the convenience of the reader, we have collected in Section II the main properties of the regularization process $\varphi \to \varphi_r$.

Suppose now we are faced with a d.c. function (a difference of convex functions) $f = g - h$, with g and h in $\Gamma_0(H)$. *How can we regularize* f? Of course, one could apply some regularization process from nonconvex analysis directly to f (from [11] for example), but, in doing so, one would pass over the specific structure in the data, namely *the presence of convexity twice* (both g and h are convex). A straightforward idea is then to regularize g and h separately and to see what happens for $f_r = g \,\square\, r/2 \|\cdot\|^2 - h \,\square\, r/2 \|\cdot\|^2$. This way, sketched in several places ([8, 10, 16], has not been developed in full yet, except for a consistent contribution in that direction by Gabay [10] (a paper which deserved to be better known) where it plays a preparatory role for designing algorithmic procedures for finding a critical point of a d.c. function. We present in Section III the main fundamental properties of f_r with respect to those of f, especially those concerning the comparison of critical points of f and f_r and the behavior of critical points of f_r when $r \to +\infty$. It thus is shown that, to a great extent, the nice properties of the regularization process $\varphi \to \varphi \,\square\, r/2 \|\cdot\|^2$ when applied to convex functions φ are preserved for the process $f \to f_r$ performed on d.c. functions f.

One could also make use of other kernel functions like $r \|\cdot\|$, which actually offers some advantages over the regularization process with

$r/2\ \|\cdot\|^2$ (cf. [15]). The regularization of a d.c. function with a general kernel function $\theta \in \Gamma_0(H)$ has been studied in a recent work by Plazanet [22], but the kernel function $r/2\ \|\cdot\|^2$ brings in peculiar properties which deserve to be presented in a distinct and clear-cut manner.

II. Regularizing a Convex Function with the Kernel $r/2\ \|\cdot\|^2$.

Given $\varphi \in \Gamma_0(H)$ and $r > 0$, the *infimal convolution* of φ with the *kernel function* $r/2\ \|\cdot\|^2$ is the function φ_r denoted as $\varphi \,\Box\, r/2\ \|\cdot\|^2$ and defined on H as

$$\forall x \in H, \qquad \varphi_r(x) = \inf_{u_1 + u_2 = x} \left\{ \varphi(u_1) + \frac{r}{2} \|u_2\|^2 \right\} \tag{2.1}$$

$$= \inf_{u \in H} \left\{ \varphi(u) + \frac{r}{2} \|x - u\|^2 \right\}. \tag{2.2}$$

"Convolution" stems from the fact that the operation $\Box$ has a smoothing (or regularizing) effect when it is performed with a differentiable convex function, somewhat like the usual (integral) convolution $*$ in analysis. "Infimal" reminds us of the infimum taken in the definition of $\varphi_r(x)$. Geometrically, performing φ_r amounts to adding two convex sets in $H \times \mathbb{R}$; more precisely the strict epigraph of φ_r (i.e., the set of (x, α) for which $\varphi_r(x) < \alpha$) is the sum of the strict epigraphs of φ and $r/2\ \|\cdot\|^2$. This has led some authors [5] to call "epigraphic addition" what is traditionally called infimal convolution. The functions φ_r, $r > 0$, are often also called Moreau–Yosida approximates of φ.

Due to the nice properties of the kernel function $r/2\ \|\cdot\|^2$, the function φ_r then enjoys peculiar properties, widely known in the field of convex analysis (see [13, 6, 2, 1, pp. 21–24 and 66–68] for instance), which we recall now.

Repeating the process of regularization does not add anything new since

$$(\mathscr{P}_0) \qquad (\varphi_r)_{r'} = \varphi_{r''}, \qquad \text{where } 1/r'' = 1/r + 1/r'.$$

This can easily be checked by playing with the conjugacy operation

$$(\varphi_r)^* = \varphi^* + \frac{1}{2r} \|\cdot\|^2 \qquad \text{and} \qquad (\varphi_r^*)^* = \varphi_r.$$

For a fixed x, the function $r \to \varphi_r(x)$ is locally Lipschitz [2] and its behavior when $r \to 0^+$ and $r \to +\infty$ is checked later.

($\mathscr{P}_1$) The infimum in the definition (2.2) of $\varphi_r(x)$ is achieved at a unique point, denoted x_r^φ, and characterized by the equation

$$r(x - x_r^\varphi) \in \partial\varphi(x_r^\varphi). \tag{2.3}$$

This relation is nothing other than the optimality condition in the minimization problem defining $\varphi_r(x)$. By developing it we get that $x \in (I+(1/r)\,\partial\varphi)(x_r^\varphi)$ (where $\Gamma : H \to H$ denotes the identity mapping); what we have claimed above says that the set-valued mapping $I+(1/r)\,\partial\varphi : H \to H$ is surjective and

$$\left(I+\frac{1}{r}\,\partial\varphi\right)^{-1} : x \to \left(I+\frac{1}{r}\,\partial\varphi\right)^{-1}(x) = x_r^\varphi$$

is a mapping (called the *resolvent mapping* of the maximal monotone operator $\partial\varphi$). Given x, where $\partial\varphi(x)$ is nonempty, the points whose images by the mapping $(\cdot)_r^\varphi$ is x are those of $(I+(1/r)\,\partial\varphi)$ (x). In short, $(x+x^*/r)_r^\varphi = x$ for all $x^* \in \partial\varphi(x)$.

Condition (2.3) can be expressed as a variational inequality via the definition of $\partial\varphi(x_r)$:

$$\forall y \in H, \qquad \varphi(y) - \varphi(x_r^\varphi) + r\langle x_r^\varphi - x, y - x_r^\varphi \rangle \geqslant 0. \tag{2.4}$$

But there is another variational inequality, more tricky to derive, which also characterizes x_r^φ:

$$\forall y \in H, \qquad \varphi(y) - \varphi(x_r^\varphi) + r\langle y - x, y - x_r^\varphi \rangle \geqslant 0. \tag{2.5}$$

($\mathscr{P}_2$) $\varphi_r : H \to \mathbb{R}$ is convex, Fréchet-differentiable on H with

$$\forall x \in H, \qquad \nabla\varphi_r(x) = r(x - x_r^\varphi). \tag{2.6}$$

The Gâteaux-differentiability of φ_r (which amounts to saying that $\partial\varphi_r(x)$ contains exactly one element) merely comes from the calculus rule on the subdifferential of the infimal convolution of two convex functions [12], while proving its Fréchet-differentiability requires more efforts [6].

($\mathscr{P}_3$) The mappings $x \to x_r^\varphi$ and $x \to x - x_r^\varphi$ share the same monotonicity and Lipschitz properties, namely: for all x, y in H,

$$\langle x_r^\varphi - y_r^\varphi, x - y \rangle \geqslant \| x_r^\varphi - y_r^\varphi \|^2,$$
$$\text{which implies } \| x - y \| \geqslant \| x_r^\varphi - y_r^\varphi \|; \tag{2.7}$$

$$\langle (x - x_r^\varphi) - (y - y_r^\varphi), x - y \rangle \geqslant \| (x - x_r^\varphi) - (y - y_r^\varphi) \|^2,$$
$$\text{which implies } \| x - y \| \geqslant \| (x - x_r^\varphi) - (y - y_r^\varphi) \|. \tag{2.8}$$

As a consequence, the gradient mapping $\nabla\varphi_r$ is Lipschitz on H with r as a Lipschitz constant; we say in short that φ_r is a $C^{1,1}$ *convex function.* It is then interesting to keep in mind the following bounds for the difference between $\varphi_r(y)$ and the first-order development $\varphi_r(x) + \langle \nabla\varphi_r(x), y - x\rangle$ of φ_r at x : for all x, y in H,

$$0 \leqslant \varphi_r(y) - \varphi_r(x) - r\langle x - x_r^\varphi, y - x\rangle \leqslant r \|x - y\|^2. \tag{2.9}$$

($\mathscr{P}_4$) When $r \to +\infty$, $\varphi_r(x) \to \varphi(x)$ for all $x \in H$ and $x_r^\varphi \to x$ for all $x \in \operatorname{dom} \varphi$.

(Whence $\|\nabla\varphi_r(x)\| = o(r)$ for a given $x \in \operatorname{dom}\varphi$). If H is finite-dimensional, the convergence of φ_r toward φ is uniform on each compact contained in $\widehat{\overset{o}{\operatorname{dom}}\, \varphi}$.

Concerning the behavior of $\nabla\varphi_r(x)$ when $r \to +\infty$, one can be more precise [16]: if x is a point where $\partial\varphi(x)$ is nonempty, $\nabla\varphi_r(x)$ converges toward the element of $\partial\varphi(x)$ with smallest norm.

As a general rule, $\varphi(x_r^\varphi) \leqslant \varphi_r(x) \leqslant \varphi(x)$ for all $x \in H$, but both φ_r and φ have the same lower bounds and coincide only at their minimum points:

$$\inf_{x \in H} \varphi(x) = \inf_{x \in H} \varphi_r(x). \tag{2.10}$$

The following statements are equivalent:

$$\begin{array}{ll}
\text{(i)} & x \text{ minimizes } \varphi \text{ on } H; \\
\text{(ii)} & x \text{ minimizes } \varphi_r \text{ on } H; \\
\text{(iii)} & x = x_r^\varphi; \\
\text{(iv)} & \varphi(x) = \varphi(x_r^\varphi); \\
\text{(v)} & \varphi(x) = \varphi_r(x) \ (x \text{ is a coincidence point of } \varphi \text{ and } \varphi_r).
\end{array} \tag{2.11}$$

($\mathscr{P}_5$) When $r \to 0^+$, $\varphi_r(x) \to \inf_{u \in H} \varphi(u)$. If the infimum of φ on H is achieved, then $\nabla\varphi_r(x) \to 0$ and the limit of x_r^φ (whenever it exists) minimizes φ on H.

The comparison between regularized versions of φ and φ^* is easily done thanks to Moreau's relation [14] which states that $\psi \,\square\, \frac{1}{2}\|\cdot\|^2 + \psi^* \,\square\, \frac{1}{2}\|\cdot\|^2 = \frac{1}{2}\|\cdot\|^2$ for all $\psi \in \Gamma_0(H)$. Applying this to $\psi = r\varphi(\cdot/r)$, we get immediately that

$$\varphi_r(x) + (\varphi^*)_{1/r}(rx) = \frac{r}{2}\|x\|^2 \qquad \text{for all } x \in H \tag{2.12}$$

(see [3] for another proof).

For $x \in H$, let $\varphi^* - \langle x, \cdot\rangle$ denote the function (on H) which assigns $\varphi^*(x^*) - \langle x, x^*\rangle$ to $x^* \in H$. One easily checks that $(\varphi^* - \langle x, \cdot\rangle)_{1/r}(0) = (\varphi^*)_{1/r}(rx) - r/2\|x\|^2$. We then have

($\mathscr{P}_6$) $\varphi_r(x) = -(\varphi^* - \langle x, \cdot\rangle)_{1/r}(0) = r/2\|x\|^2 - (\varphi^*)_{1/r}(rx)$, and the solution to the minimization problem defining $(\varphi^* - \langle x, \cdot\rangle_{1/r}(0)$ or $(\varphi^*)_{1/r}(rx)$ is precisely $\nabla\varphi_r(x) = r(x - x_r^\varphi)$:

Note the particular role played by $x=0$ since

$$\varphi_r(0)+(\varphi^*)_{1/r}(0)=0 \qquad \text{for all } r>0. \tag{2.13}$$

III. Regularizing a d.c. Function with the Kernel $r/2\,\|\cdot\|^2$

Let f be a d.c. function on H, i.e., $f=g-h$ with both g and h in $\Gamma_0(H)$. Since our interest is in minimizing f, we set $f(x)=+\infty$ whenever $x\notin \operatorname{dom} g \cup \operatorname{dom} h$. In most applications however, h is (finite and) continuous on H so that no ambiguity arises in the definition of f. For $r>0$ we set

$$f_r=g_r-h_r=g \,\square\, \frac{r}{2}\|\cdot\|^2-h\,\square\,\frac{r}{2}\|\cdot\|^2.$$

This notation should not mislead the reader, but f_r is not $f\,\square\, r/2\,\|\cdot\|^2$. Note also that f_r is not the regularized version of f according to a procedure proposed by Lasry and Lions [11] and further developed by Attouch and Azé [4].

The new function f_r is defined as the difference of the regularized versions of g and h; it therefore depends on the decomposition of f as $g-h$.

III.1. We begin by casting the properties of f_r directly derived from those of g_r and h_r (cf. Section II).

(Q_1) $f_r: H\to\mathbb{R}$ is a d.c. function, Fréchet-differentiable on H with

$$\forall x\in H, \qquad \nabla f_r(x)=-r(x_r^g-x_r^h). \tag{3.1}$$

Moreover the inequality (2.9) induces that for all x, y in H,

$$|f_r(y)-f_r(x)-\langle\nabla f_r(x), y-x\rangle|\leqslant r\,\|y-x\|^2. \tag{3.2}$$

Clearly f_r is a $C^{1,1}$ d.c. function on H (∇f_r is a Lipschitz mapping on H with $2r$ as a Lipschitz constant). But one can say more: f_r is "not too far from a convex function." Indeed, we derive from (3.2),

$$|\langle\nabla f_r(y)-\nabla f_r(x), y-x\rangle|\leqslant 2r\,\|x-y\|^2 \text{ for all } x, y \text{ in } H,$$

whence ∇f_r+2rI is a monotone mapping on H and

(Q_2) $f_r+r\,\|\cdot\|^2$ is a convex function on H.

(Q_3) When $r\to+\infty$, $f_r(x)\to f(x)$ for all $x\in\operatorname{dom} g\cup\operatorname{dom} h$ and $\|\nabla f_r(x)\|=o(r)$ for all $x\in\operatorname{dom} g\cap\operatorname{dom} h$. If moreover H is finite-dimensional, the convergence of f_r toward f is uniform on each compact contained in $\overset{o}{\widehat{\operatorname{dom} g}}\cap\overset{o}{\widehat{\operatorname{dom} h}}$. If both $\partial g(x)$ and $\partial h(x)$ are nonempty, $\nabla f_r(x)$ converges toward an element of $\partial g(x)-\partial h(x)$.

(Q_4) *Bounds for* $f_r(x)$. We have for all $x \in H$

$$g(x_r^g) - h(x) \leqslant f_r(x) \leqslant g(x) - h(x_r^h). \tag{3.3}$$

(Q_5) If one of the functions g or h is bounded from below $f_r(x) \to \inf_{u \in H} g(u) - \inf_{u \in H} h(u)$ when $r \to 0^+$. If moreover the infima of both g and h on H are achieved, then $\lim_{r \to 0^+} \nabla f_r(x) = 0$.

It has been observed for a long time [24, 25] that it is very useful to associate $f^\diamond = h^* - g^*$ with $f = g - h$, especially as $\inf_{x \in H} f(x) = \inf_{x^* \in H} f^\diamond(x^*)$. Now, since $(h_r)^* = h^* + 1/2r \, \|\cdot\|^2$ and $(g_r)^* = g^* + 1/2r \, \|\cdot\|^2$, one immediately gets that

$$(f_r)^\diamond = f^\diamond \qquad \text{and} \qquad \inf_{x \in H} f(x) = \inf_{x \in H} f_r(x). \tag{3.4}$$

Now, in view of ($\mathscr{P}_6$), we have that

$$\begin{aligned} f_r(x) = g_r(x) - h_r(x) &= (h^* - \langle x, \cdot \rangle)_{1/r}(0) - (g^* - \langle x, \cdot \rangle)_{1/r}(0) \\ &= (h^*)_{1/r}(rx) - (g^*)_{1/r}(rx). \end{aligned} \tag{3.5}$$

So, the regularization process $(\cdot)_{1/r}$ performed on $f^\diamond = h^* - g^*$ boils down (on condition of a change of scaling) to f_r. The situation with respect to the $\diamond$-involution and the $(\cdot)_s$ regularization process is summarized in the diagram below.

$$\begin{tikzcd}
\cdot \arrow[ddd, "(\cdot)_s"'] & f = g - h \arrow[r, "\diamond"] \arrow[d, "(\cdot)_r"] & f^\diamond = h^* - g^* \arrow[d, "(\cdot)_{1/r}"] \\
& f_r = g_r - h_r \arrow[r, equal] \arrow[dd, "(\cdot)_{r'}"] & (f^\diamond)_{1/r}(r\cdot) = (h^*)_{1/r}(r\cdot) - (g^*)_{1/r}(r\cdot) \\
& & \\
\cdot & (f_r)_{r'} = f_{r''} = g_{r''} - h_{r''} & \\
& \text{with } 1/r'' = 1/r + 1/r' &
\end{tikzcd}$$

III.2. *Comparison of Critical Points of f and f_r.* Since f_r is a differentiable function, defining the notation of critical point of f_r does not raise any particular question: x is a critical point of f_r if and only if $\nabla f_r(x) = 0$, which amounts to $x_r^g = x_r^h$. So, critical points of f_r are those x at which the resolvent mappings of ∂g and ∂h coincide. The question is different for $f = g - h$. What definition should be considered as critical points of f? For various reasons—especially the nice correspondance between critical points $f = g - h$ and those of $f^\diamond = h^* - g^*$, we adopt the definition of critical point of f in Toland's sense [24, 25]: $x \in H$ is called a T-critical point of $f = g - h$ if $\partial g(x) \cap \partial h(x) \neq \varnothing$ (or, equivalently, $0 \in \partial g(x) - \partial h(x)$). When x is a local minimum of f, then $\partial h(x) \subset \partial g(x)$ and x is a T-critical point of f provided that $\partial h(x) \neq \emptyset$. We recall for the sake of completeness that a necessary and sufficient condition for x being a global minimum of f on H is: $\partial_\varepsilon h(x) \subset \partial_\varepsilon g(x)$ for all $\varepsilon > 0$ [18].

In view of property $(\mathscr{P}_6)$ and relation (3.5), we clearly have

$$\begin{pmatrix} x \text{ critical point} \\ \text{of } f_r = g_r - h_r \end{pmatrix} \Leftrightarrow \begin{pmatrix} 0 \text{ critical point} \\ \text{of } (h^* - \langle x, \cdot \rangle)_{1/r} - (g^* - \langle x, \cdot \rangle)_{1/r} \end{pmatrix}$$
$$\Leftrightarrow \begin{pmatrix} rx \text{ critical point} \\ \text{of } (h^*)_{1/r} - (g^*)_{1/r} \end{pmatrix}.$$

When x is a critical point of f_r, we just write x_r for $x_r^g = x_r^h$.

The first point is to compare critical points of f_r with T-critical points of f.

THEOREM 3.1. *Let x be a critical point of $f_r = g_r - h_r$. Then*:

(i) *x_r is a T-critical point of $f = g - h$ and the critical values are equal, i.e., $f_r(x) = f(x_r)$.*

(ii) *$x_r^\diamond = r(x - x_r)$ is a T-critical point of $f^\diamond = h^* - g^*$ and $f(x_r) = f^\diamond(x_r^\diamond)$.*

Proof. (i) Since x_r is the common vector $x_r^g = x_r^h$, we have

$$r(x - x_r) \in \partial g(x_r) \qquad \text{and} \qquad r(x - x_r) \in \partial h(x_r), \tag{3.6}$$

whence $\partial g(x_r) \cap \partial h(x_r) \neq \emptyset$ and x_r is a T-critical point of $f = g - h$. Moreover, (3.6) induces that

$$x_r \in \partial g^*[r(x - x_r)] \qquad \text{and} \qquad x_r \in \partial h^*[r(x - x_r)],$$

which means that $r(x - x_r)$ is a T-critical point of $f^\diamond = h^* - g^*$. Now, since

$$g_r(x) = g(x_r) + \frac{r}{2} \|x - x_r\|^2 \qquad \text{and} \qquad h_r(x) = h(x_r) + \frac{r}{2} \|x - x_r\|^2,$$

one gets that $f_r(x) = g_r(x) - h_r(x) = g(x_r) - h(x_r) = f(x_r)$.

(ii) If x_r is a T-critical point of $f = g - h$, any element x^* in $\partial g(x_r) \cap \partial h(x_r)$ is a T-critical point of $f^\diamond = h^* - g^*$ and $f^\diamond(x^*) = f(x_r)$ [7, 25]. So we just apply this to $x^* = x_r^\diamond$. ∎

Remark 3.2. To have a critical point x of f_r equal to x_r is an exceptional situation. That would mean that x minimizes g *and* h on H and

- x is a T-critical point of $f = g - h$,

- 0 is a T-critical point of $f^{\diamond} = h^* - g^*$,
- $f_r(x) = f(x) = f^{\diamond}(0) = \inf_{x \in H} g(x) - \inf_{x \in H} h(x)$.

An example illustrating such a situation is shown later.

If the minimum of f_r on H is achieved at some x, x is indeed a critical point of f_r. A natural question which arises now is: Does the corresponding T-critical point x_r of $f = g - h$ minimize f on H? The answer is yes and detailed in the next proposition.

PROPOSITION 3.3. *Let x minimize $f_r = g_r - h_r$ on H. Then x_r minimizes $f = g - h$ on H, $x_r^{\diamond} = r(x - x_r)$ minimizes $f^{\diamond} = h^* - g^*$ on H and*

$$f_r(x) = \inf_H f_r = f(x_r) = \inf_H f = f^{\diamond}(x_r^{\diamond}) = \inf_H f^{\diamond}. \tag{3.7}$$

Note incidentally that in the situation described in the proposition, $\partial h(x_r)$ is necessarily nonempty (because x_r is a T-critical point of $f = g - h$) and $\partial h(x_r) \subset \partial g(x_r)$. Also note that rx minimizes $(h^*)_{1/r} - (g^*)_{1/r}$ on H, and $(h^*)_{1/r}\,(rx) - (g^*)_{1/r}\,(rx)$ equals the common value (3.7).

Proof. It suffices to combine the results of Theorem 3.1 with the following string of equalities (cf. (3.4))

$$\inf_H f_r = \inf_H f = \inf_H f^{\diamond} \quad \blacksquare$$

According to Theorem 3.1 and Proposition 3.3, if f_r has critical points (resp. minimum points), then $f = g - h$ has T-critical points (resp. minimum points). This is viewed through the transformation $(\cdot)_r : x \to x_r$;

$$\left\{x \,\middle|\, \begin{array}{c} x \text{ critical} \\ \text{points of } f_r \end{array}\right\} \xrightarrow{(\cdot)_r} \left\{x_r \,\middle|\, \begin{array}{c} x \text{ critical} \\ \text{points of } f \end{array}\right\} \subset \left\{\begin{array}{c} T\text{-critical} \\ \text{points of } f_r \end{array}\right\}. \tag{3.8}$$

The converse correspondence is explored now: given a T-critical point of f, is it the image by $(\cdot)_r$ of some critical point of f_r?

THEOREM 3.4. *Let x be a T-critical point of $f = g - h$. Then, for all $x^* \in \partial g(x) \cap \partial h(x)$, $x + (x^*/r)$ is a critical point of $f_r = g_r - h_r$ and the critical values are equal, i.e., $f(x) = f_r\,(x + x^*/r)$.*

If, moreover, x is a minimizer of $f = g - h$ on H, then $x + x^*/r$ is a minimizer of $f_r = g_r - h_r$ on H and

$$f(x) = \inf_H f = f_r\left(x + \frac{x^*}{r}\right) = \inf_H f_r.$$

Note, in addition, some consequences we already observed : x^* is a T-critical point of $f^\diamond = h^* - g^*$, $f(x) = f^\diamond(x^*)$, and $x^* + rx$ is a critical point of $(h^*)_{1/r} - (g^*)_{1/r}$.

In view of the theorem above, any T-critical point of f is achieved through the transformation $(\cdot)_r$: in short, if x is a T-critical point of f and $x^* \in \partial g(x) \cap \partial h(x)$, we have $(x + x^*/r)_r = x$.

Proof. Since $x^* \in \partial g(x) \cap \partial h(x)$,

$$x + \frac{x^*}{r} \in \left(I + \frac{1}{r}\,\partial g\right)(x) \cap \left(I + \frac{1}{r}\,\partial h\right)(x).$$

whence

$$\left(x + \frac{x^*}{r}\right)_r^g = \left(x + \frac{x^*}{r}\right)_r^h = x.$$

The rest follows as in the proofs of Theorem 3.1 and Proposition 3.3. ∎

As we recalled in Section II, for $\varphi \in \Gamma_0(H)$ all the x' giving rise to the same x_r^φ are those of $(I + (1/r)\,\partial\varphi)(x_r^\varphi)$. So, any critical point of $f_r = g_r - h_r$ is of the form $x + x^*/r$, with x a T-critical point of $f = g - h$ and $x^* \in \partial g(x) \cap \partial h(x)$. As a result: *when $r \to +\infty$, the critical points (resp. the minimizers) of $f_r = g_r - h_r$ converge to T-critical points (resp. minimizers) of $f = g - h$ and all T-critical points (resp. minimizers which are T-critical points) of $f = g - h$ are attained in such a way.*

A particular case with respect to minimization is the following (as already observed by Gabay [10]): assume x is the unique minimizer T-critical point of $f = g - h$ and x^* is the only minimizer of $f^\diamond = h^* - g^*$, then $x + x^*/r$ is the only minimizer of $f_r = g_r - h_r$.

All the properties described in Theorem 3.1, Proposition 3.3, and Theorem 3.4 are illustrated in the following simple examples.

EXAMPLE 3.5. Let $g : x \to g(x) = \frac{1}{2}x^2$, $h : x \to h(x) = |x|$ and $f(x) = g(x) - h(x)$. The T-critical points of f are -1, 0, and $+1$, with

$$f(-1) = f(+1) = \inf_{\mathbb{R}} f.$$

The regularized versions of g and h—hence of f—are described as

$$\forall x \in \mathbb{R}, \qquad x_r^g = \frac{r}{r+1}\,x \qquad \text{and} \qquad g_r(x) = \frac{1}{2}\,\frac{r}{r+1}\,x^2;$$

$$x_r^h = \begin{cases} 0 \text{ if } |x| \leqslant 1/r \\ x - 1/r \text{ if } x \geqslant 1/r \\ x + 1/r \text{ if } x \leqslant -1/r \end{cases} \qquad \text{and} \qquad h_r(x) = \begin{cases} r/2x^2 \text{ if } |x| \leqslant 1/r, \\ |x| - 1/2r \text{ if } |x| \geqslant 1/r \end{cases}$$

$$f_r(x)=\begin{cases} -\dfrac{r^2}{2(r+1)}x^2 \text{ if } |x|\leqslant 1/r, \\ \dfrac{1}{2}\dfrac{r}{r+1}x^2-|x|+1/2r \text{ if } |x|\geqslant 1/r. \end{cases}$$

Thus, the critical points of f_r are $-1-1/r$, 0 and $1+1/r$, with

$$f_r\left(-1-\frac{1}{r}\right)=f_r\left(1+\frac{1}{r}\right)=\inf_{\mathbb{R}} f_r.$$

In this example, $x=0$ is the only critical point where $x=x_r$ (cf. Remark following Theorem 3.1).

The "dual" function $f^\diamond=h^*-g^*$ and its regularized versions are easy to calculate here. We have

$$\forall x\in\mathbb{R},\qquad h^*(x)=0 \qquad \text{if} \quad |x|\leqslant 1, \qquad +\infty \text{ if not}; g^*=g;$$

$$(h^*)_s(x)=0 \qquad \text{if} \quad |x|\leqslant 1, \qquad \frac{s}{2}(|x|-1)^2 \text{ if not};$$

$$(f^\diamond)_s(x)=(h^*)_s(x)-(g^*)_s(x)=\begin{cases} -\dfrac{1}{2}\dfrac{s}{s+1}x^2 & \text{if} \quad |x|\leqslant 1 \\ \dfrac{1}{2}\dfrac{s^2}{s+1}x^2-s|x|+\dfrac{s}{2} & \text{if} \quad |x|\geqslant 1. \end{cases}$$

The critical points of $(f^\diamond)_s$ are $-1-1/s$, 0, $1+1/s$, while the T-critical points of $f^\diamond$ are -1, 0, and $+1$.

EXAMPLE 3.6. Let $g: x\to g(x)=0$ if $|x|\leqslant 1$, $+\infty$ if not, $h: x\to h(x)=-(1-x^2)^{1/2}$ if $|x|\leqslant 1$, $+\infty$ if not. The resulting $f=g-h$ takes the value $+\infty$ outside of $[-1,+1]$, and although -1 and $+1$ minimize f on $\mathbb{R}$, they are not T-critical points of f.

Although h_r and the x_r^h cannot be calculated explicitly, we know without going further that $x=0$ is the only critical point of $f_r=g_r-h_r$ (0 is again a case where $x=x_r$) and that the infimum of f_r (equal to 0) *is not achieved.*

EXAMPLE 3.7. Let C be a nonempty closed convex set of H and $h: H\to\mathbb{R}$ a convex function. We consider the problem of maximizing $h(x)$ over C. Since

$$\sup_{x\in C} h(x)=-\inf_{x\in H}\{\psi_C(x)-h(x)\},$$

where $\psi_C(\in \Gamma_0(H))$ denotes the indicator function of C, the $x \in C$ maximizing $h(x)$ over C are those minimizing $\psi_C(x) - h(x)$ over H. The T-critical points in the original problem are those $x \in C$ for which $\partial h(x) \cap N(C; x)$ is nonempty, where $N(C; x)$ stands for the normal cone to C at x. According to our methodology, a way of approximating these T-critical points is to consider critical points of the regularized version f_r of $f = \psi_C - h$. Here $f_r = (r/2)\, d_C^2 - h_r$, where d_C denotes the distance function to the set C, and x is a critical point of f_r if and only if x_r^h is the projection of x on C.

IV. Final Remarks and Conclusion

It is tempting to extend the above displayed results on critical points to "approximate critical points" defined via ε-subdifferentials, $\varepsilon > 0$, instead of (exact) subdifferentials. The indeed can be done as far as the comparison of approximate T-critical points of $f = g - h$ and $f^\diamond = h^* - g^*$ is concerned; things are less pleasant when the objective is to associate approximate critical points of $f_r = g_r - h_r$ with approximate T-critical points of $f = g - h$ [17]. Let us mention some results in that respect. Given $\varepsilon > 0$, x is said to be an ε-T-critical point of $f = g - h$ if $\partial_\varepsilon g(x) \cap \partial_\varepsilon h(x)$ is nonempty. We then have:

(i) If x is an ε-T-critical point of $f = g - h$ and $x^* \in \partial_\varepsilon g(x) \cap \partial_\varepsilon h(x)$, then x^* is an ε-T-critical point of $f^\diamond = h^* - g^*$ and $|f(x) - f^\diamond(x^*)| \leqslant \varepsilon$.

(ii) Estimate of the distance between ε-T-critical values: if x_1 and x_2 are T-critical points of $f = g - h$, and $x_1^* \in \partial_\varepsilon g(x_1) \cap \partial_\varepsilon h(x_2)$, $x_2^* \in \partial_\varepsilon g(x_2) \cap \partial_\varepsilon h(x_2)$, then $|f(x_1) - f(x_2)| \leqslant \langle x_2^* - x_1^*, x_2 - x_1 \rangle + 2\varepsilon$.

(iii) If x_α minimizes $f = g - h$ on H within α and $x_{\alpha,\varepsilon}^* \in \partial_\varepsilon h(x_\alpha)$, then $x_{\alpha,\varepsilon}^*$ minimizes $f^\diamond = h^* - g^*$ within $\alpha + \varepsilon$.

As we said earlier, results associating ε-critical points of $f_r = g_r - h_r$ (i.e., those for which $\partial_\varepsilon g_r(x) \cap \partial_\varepsilon h_r(x) \neq \varnothing$) with ε-T-critical points of $f = g - h$ are not as complete and pleasant as those for the "limiting case" $\varepsilon = 0$.

Algorithmic procedure using regularized version f_r of f for finding T-critical points of f are not fully explored up to now, even if substantial progress has been made in some particular situations [20, 21, 23]. We think that analogues of "proximal methods" in convex optimization could be designed for d.c. optimization, by performing for example one step of a proximal method on g followed by another step of a proximal method on h, so that the presence of convexity twice in $f = g - h$ is taken into account. However, sequences generated by such potential methods are expected to converge to T-critical points of f only and not toward (global) minimizers of f. Algorithms for finding global minimizers of d.c. functions should incorporate in their construction informations from necessary and sufficient conditions for global optimality [18].

Addenda (October 1989). B. Lemaire (University of Montpellier) has just informed us of a recent work he has done on proximal methods for d.c. minimization precisely (Ref. [20]).

References

1. J.-P. Aubin, "L'analyse non linéaire et ses motivations économiques," Masson, Paris, 1984.
2. H. Attouch, "Variational Convergence for Functions and Operators," Pitman, London, 1984.
3. H. Attouch and R. Wets, Isometries for the Legendre–Fenchel transform, *Trans. Amer. Math. Soc.* **296** (1986), 33–60.
4. H. Attouch and D. Aze, Regularization and approximation of sets and functions in Hilbert spaces, *in* "Séminaire d'Analyse Numérique," Paper XI, Université Paul Sabatier de Toulouse, 1987/1988.
5. H. Attouch and R. Wets, Epigraphic analysis, *in* "Analyse non linéaire," pp. 73–99, Gauthiers-Villars, Paris, 1989.
6. H. Brezis, "Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert," North-Holland, Amsterdam, 1973.
7. I. Ekeland and T. Turnbull, Infinite-dimensional optimization and convexity, *in* "Chicago Lectures in Mathematics Series," Chicago, 1983.
8. R. Ellaia, "Contribution à l'analyse et l'optimisation de différences de fonctions convexes," Thesis, L'Université Paul Sabatier, Toulouse, 1984.
9. E. Fernandez-Cara and C. Moreno, Critical point approximation through exact regularization, *Math. Comput.* **50** (1988), 139–153.
10. D. Gabay, Minimizing the difference of two convex functions. I. Algorithms based on exact regularization, Working paper, INRIA, 1982.
11. J.-M. Lasry and P.-L. Lions, A remark on regularization in Hilbert spaces, *Israel J. Math.* **55**, No. 3 (1986), 257–266.
12. P.-J. Laurent, "Approximation et optimisation," Hermann, Paris, 1972.
13. J.-J. Moreau, Proximité et dualité dans un espace hilbertien, *Bull. Soc. Math. France* **93** (1965), 273–299.
14. J.-J. Moreau, Weak and strong solutions of dual problems, *in* "Contributions to Non-linear Functional Analysis" (Eduardo H. Zarantonello, Ed.), Academic Press, New York, 1971.
15. J.-B. Hiriart-Urruty, Lipschitz r-continuity of the approximate subdifferential of a convex function, *Math. Scand.* **47** (1980), 123–134.
16. J.-B. Hiriart-Urruty, Generalized differentiability, duality and optimization for problems dealing with differences of convex functions, *in* "Convexity and Duality in Optimization," Lecture Notes in Economics and Mathematical Systems, Vol. 256, pp. 37–70, 1986.
17. J.-B. Hiriart-Urruty, Extrémalité et dualité en optimisation non convexe, Cours de D.E.A., Université Paul Sabatier, Toulouse, 1988/1989.
18. J.-B. Hiriart-Urruty, From convex optimization to nonconvex optimization: Necessary and sufficient conditions for global optimality, *in* "Nonsmooth Optimization and Related Topics," pp. 219–240, Plenum, New York/London, 1989.

19. B. LEMAIRE, The proximal algorithm, *in* "New methods in Optimization and Their Industrial Uses," International Series of Numerical Mathematics, Vol. 87, pp. 73–87, Birkhäuser, Basel, 1989.
20. B. LEMAIRE, Proximal methods in d.c. optimization, *in* "Travaux du Séminaire d'Analyse Convexe, Montpellier, 1990," to appear.
21. PHAM DINH T. AND EL BERNOUSSI S., Duality in d.c. (difference of convex functions) optimization: Subgradient methods, *in* "Trends in Mathematical Optimization" (Hoffman *et al.*, Eds.), International Series of Numerical Mathematics, Vol. 84, pp. 277–293, Birkhäuser, Basel, 1988.
22. PH. PLAZANET, preprint Université Paul Sabatier, October 1989.
23. R. SAMPAIO, personal communication, Rio de Janeiro, October 1988.
24. J. TOLAND, Duality in nonconvex optimization, *J. Math. Anal. Appl.* **66** (1978), 399–415.
25. J. TOLAND, A duality principle for nonconvex optimization and the calculus of variations, *Arch. Rational Mech. Anal.* **71** (1979), 41–61.